

# Mimicking enzyme evolution by generating new $(\beta\alpha)_8$ -barrels from $(\beta\alpha)_4$ -half-barrels

Birte Höcker<sup>†‡</sup>, Jörg Claren<sup>‡§</sup>, and Reinhard Sterner<sup>†§¶</sup>

<sup>†</sup>Institut für Biochemie, Universität zu Köln, Otto-Fischer-Strasse 12-14, D-50674 Köln, Germany; and <sup>§</sup>Institut für Biophysik und Physikalische Biochemie, Universität Regensburg, Universitätsstrasse 31, D-93053 Regensburg, Germany

Edited by Gregory A. Petsko, Brandeis University, Waltham, MA, and approved October 5, 2004 (received for review August 9, 2004)

Gene duplication and fusion events that multiply and link functional protein domains are crucial mechanisms of enzyme evolution. The analysis of amino acid sequences and three-dimensional structures suggested that the  $(\beta\alpha)_8$ -barrel, which is the most frequent fold among enzymes, has evolved by the duplication, fusion, and mixing of  $(\beta\alpha)_4$ -half-barrel domains. Here, we mimicked this evolutionary strategy by generating *in vitro*  $(\beta\alpha)_8$ -barrels from  $(\beta\alpha)_4$ -half-barrels that were deduced from the enzymes imidazole glycerol phosphate synthase (HisF) and *N*'[(5'-phosphoribosyl)formimino]-5-aminoimidazole-4-carboxamide-ribonucleotide isomerase (HisA). To this end, the gene for the C-terminal  $(\beta\alpha)_4$ -half-barrel (HisF-C) of HisF was duplicated and fused in tandem to yield HisF-CC, which is more stable than HisF-C. In the next step, by optimizing side-chain interactions within the center of the  $\beta$ -barrel of HisF-CC, the monomeric and compact  $(\beta\alpha)_8$ -barrel protein HisF-C\**C* was generated. Moreover, the genes for the N- and C-terminal  $(\beta\alpha)_4$ -half-barrels of HisF and HisA were fused crosswise to yield the chimeric proteins HisFA and HisAF. Whereas HisFA contains native secondary structure elements but adopts ill-defined association states, the  $(\beta\alpha)_8$ -barrel HisAF is a stable and compact monomer that reversibly unfolds with high cooperativity. The results obtained suggest a previously undescribed dimension for the diversification of enzymatic activities: new  $(\beta\alpha)_8$ -barrels with novel functions might have evolved by the exchange of  $(\beta\alpha)_4$ -half-barrel domains with distinct functional properties.

chimeric proteins | gene duplication | histidine biosynthesis | TIM-barrel

Gene duplication plays an important role in enzyme evolution. It has been estimated that  $\approx 50\%$  of all genes in microorganisms are the result of duplication events, which are followed by diversification of the twin genes (1, 2). Because the  $(\beta\alpha)_8$ - or TIM-barrel is the most common structural scaffold among enzymes, its duplication must have occurred frequently. The fold of the canonical  $(\beta\alpha)_8$ -barrel consists of a central barrel of eight parallel  $\beta$ -strands and eight external  $\alpha$ -helices. Connecting loops are located at the N-terminal ( $\alpha$ - $\beta$  loops) and C-terminal ( $\beta$ - $\alpha$  loops) face of the barrel (3, 4). Although  $(\beta\alpha)_8$ -barrel enzymes catalyze a broad range of chemically diverse reactions, their active sites always are located at the C-terminal face of the barrel (5). For this reason, and on the basis of comprehensive amino acid sequence comparisons (6, 7), it has been postulated that a large fraction of the known  $(\beta\alpha)_8$ -barrels have evolved by gene duplication and diversification (8, 9). In particular, similarities in sequence, structure, and function suggest that several  $(\beta\alpha)_8$ -barrels from the tryptophan and histidine biosynthetic pathways have evolved divergently from a common ancestral enzyme (10, 11). In support of this hypothesis, both *N*'[(5'-phosphoribosyl)formimino]-5-aminoimidazole-4-carboxamide-ribonucleotide (ProFAR) isomerase (HisA) and imidazole glycerol phosphate synthase (HisF), which catalyze two consecutive reactions of histidine biosynthesis, could be converted by exchanging a single amino acid into enzymes with phosphoribosylanthranilate isomerase (TrpF) activity (12, 13).

$(\beta\alpha)_8$ -Barrels are considered to be single-domain proteins. HisF and HisA from *Thermotoga maritima* (Fig. 1) possess,

however, a striking internal twofold symmetry. The pairs of N-terminal halves (designated HisF-N and HisA-N), which consist of the first four  $(\beta\alpha)$  units, and the pairs of C-terminal halves (designated HisF-C and HisA-C), which consist of the last four  $(\beta\alpha)$  units, display sequence identities between 16% and 26% and rms deviation values of their main-chain nonhydrogen atoms between 1.4 and 2.1 Å (10, 14). Moreover, the catalytically essential aspartate residues of both HisF and HisA are located at equivalent positions at the C-terminal ends of the respective strands  $\beta 1$  (within HisF-N and HisA-N) and  $\beta 5$  (within HisF-C and HisA-C) (10, 15–17). When produced separately, the half-barrels HisF-N and HisF-C are homodimeric proteins with native secondary and tertiary structures but without measurable catalytic activity. When coexpressed *in vivo* or refolded together *in vitro*, the two proteins assemble to a catalytically fully active HisF-NC heterodimer (15). It appears that both HisF and HisA are composed of two structural domains, namely the corresponding N- and C-terminal half-barrels. These results suggest an evolutionary scenario according to which a primordial gene encoding a  $(\beta\alpha)_4$ -half-barrel as a subunit of a homodimeric enzyme was duplicated and fused to yield a monomeric, ancestral  $(\beta\alpha)_8$ -barrel, from which HisF, HisA, and presumably also TrpF evolved by a series of further gene duplication and diversification events (4, 11). Moreover, it was postulated that  $(\beta\alpha)_4$ -half-barrels are independently evolving domains, implying that new  $(\beta\alpha)_8$ -barrels could be generated by mixing and joining  $(\beta\alpha)_4$ -half-barrels from an existing pool (18). Along these lines, an extensive search of the Protein Data Bank ([www.pdb.org](http://www.pdb.org)) with HisF-N and HisF-C as queries revealed significant similarities to several  $(\beta\alpha)_8$ -barrel enzymes and to members of the flavodoxin-like fold family, as indicated by sequence identities of up to 22% and Z scores of up to 6.4 (19). The flavodoxin-like fold, which is found both as an isolated polypeptide chain and as an integral domain of larger proteins, consists of five  $(\beta\alpha)$  elements. Four of them are topologically equivalent to the four  $(\beta\alpha)$  elements of HisF-N and HisF-C, and the fifth element corresponds to an additional small two-stranded  $\beta$ -sheet that is located in the first  $\beta$ - $\alpha$  loop of each half-barrel.

The goal of the present work was to reconstruct experimentally the postulated evolutionary events that can lead to the generation of new  $(\beta\alpha)_8$ -barrels from existing  $(\beta\alpha)_4$ -half-barrels. To this end, the half-barrel HisF-C was duplicated, fused, and optimized to yield the stable and monomeric HisF-C\**C* barrel. Moreover, the N- and C-terminal half-barrels of HisA and HisF were fused crosswise to yield the chimeric HisAF and HisFA proteins (Fig. 1). The results indicate that stable  $(\beta\alpha)_8$ -barrels can be assembled in the laboratory from  $(\beta\alpha)_4$ -half-barrels and

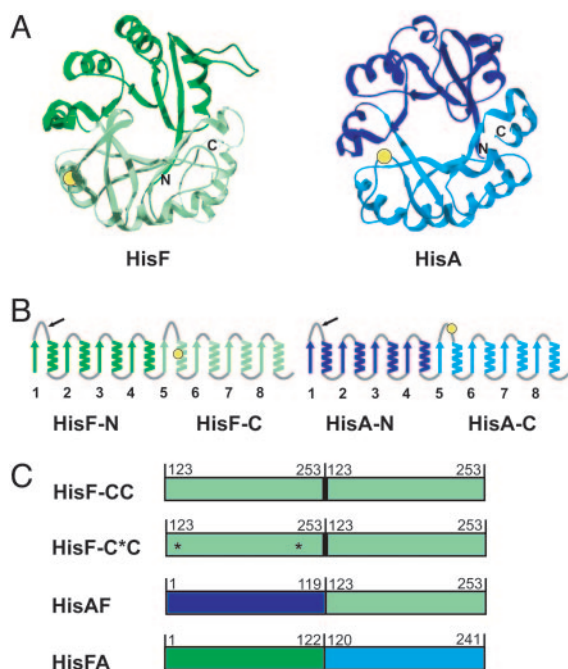
This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: HisA, *N*'[(5'-phosphoribosyl)formimino]-5-aminoimidazole-4-carboxamide-ribonucleotide isomerase; HisF, imidazole glycerol phosphate synthase.

<sup>‡</sup>Present address: Department of Biochemistry, Duke University Medical Center, Box 3711, Durham, NC 27710.

<sup>¶</sup>To whom correspondence should be addressed. E-mail: [reinhard.sterner@biologie.uni-regensburg.de](mailto:reinhard.sterner@biologie.uni-regensburg.de).

© 2004 by The National Academy of Sciences of the USA



**Fig. 1.** The (β)8-barrel enzymes HisF (green) and HisA (blue) consist of half-barrels, which were fused in different combinations. (A) Ribbon diagrams showing a view down the axis of the central β-barrels. The N-terminal half-barrels HisF-N and HisA-N are dark-colored, and the C-terminal half-barrels HisF-C and HisA-C are light-colored. The positions of the single tryptophan residues 156 (HisF) and 138 (HisA) are marked with a yellow circle. (B) Topology diagrams showing the unrolled eight (β)α units joined at the top by β-α loops and at the bottom by α-β loops. β-α loops 1 and 5 are longer and more flexible than the other loops. In both HisA and HisF, the black arrows indicate a specific trypsin cleavage site located in β-α loop 1. (C) Fusion constructs produced and characterized in this work. The residue numbers give the borders of the half-barrels of HisF and HisA. The thick bar in HisF-CC and HisF-C\*C represents the Gly-Ser-Gly linker that joins the two (β)α-half-barrels. The locations of the two amino acid exchanges at positions 124 and 220 in the N-terminal half of HisF-C\*C (see Fig. 2) are marked with asterisks.

suggest that similar events might have occurred in the course of natural evolution.

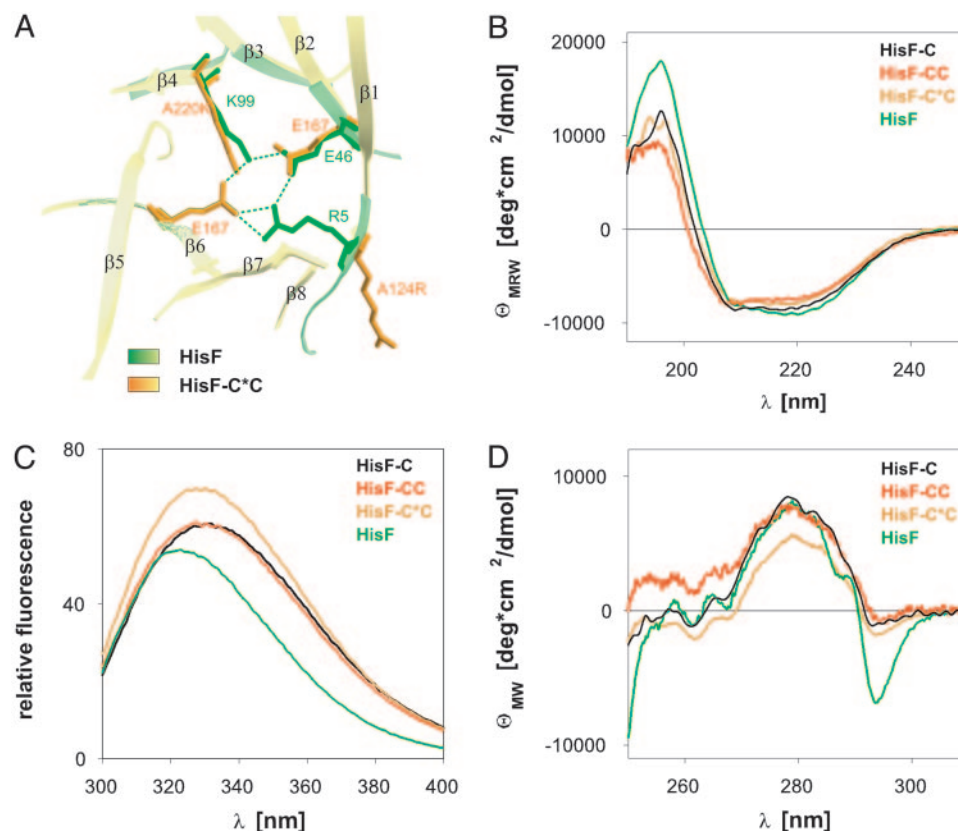
## Materials and Methods

**Cloning and Heterologous Expression of hisF-CC and hisF-C\*C and Purification of the Protein Products.** The *hisF*-CC gene was cloned in several steps. First, *hisF*-C (15) was amplified by PCR, using the plasmid SK+/III P-P as a template (14). For the amplification of *hisF*-C<sub>1</sub>, which is the 5' copy within *hisF*-CC, the oligonucleotide 5'-ATA CAT ATG CAG GCC GTT GTC GTG GCG ATA-3' with a *Nde*I site (in bold) was used as the 5' primer, and the oligonucleotide 5'-ATA GGA TCC CAA CCC CTC CAG TCT CAC GTT-3' with a *Bam*HI site (in bold) was used as the 3' primer. For the amplification of *hisF*-C<sub>2</sub>, which is the 3' copy within *hisF*-CC, the oligonucleotide 5'-ATA GGA TCC GGT CAG GCC GTT GTC GTG GCG-3' with a *Bam*HI site (in bold) was used as the 5' primer, and the oligonucleotide 5'-GTG CTC GAG CAA CCC CTC CAG TCT CAC GTT-3' with a *Xho*I site (in bold) was used as the 3' primer. The amplified *hisF*-C<sub>2</sub> was cloned into pET24a(+) by using *Bam*HI and *Xho*I, yielding the construct pET24a(+)-*hisF*-C<sub>2</sub>. Then, *hisF*-C<sub>1</sub> was cloned into this vector by using *Nde*I and *Bam*HI, yielding pET24a(+)-*hisF*-CC. The primers were designed such that a Gly-Ser-Gly linker was introduced between the two *hisF*-C units to give the fused half-barrels more conformational freedom and to avoid steric clashes that might impede their association. For the production

of *hisF*-C\*C from *hisF*-CC, the alanine codons 124 and 220 of *hisF*-C<sub>1</sub> were replaced by an arginine and a lysine codon, respectively. The mutations were introduced by megaprimer PCR (20), by using the oligonucleotide 5'-ATA CAT ATG CAG CGC GTT GTC GTG GCG ATA-3' to introduce the mutation A124R (new codon in bold), and the oligonucleotide 5'-GAC AAG GCC CTT GCG GCT TCT GTC-3' was used to introduce the mutation A220K (new codon in bold). All constructs were sequenced entirely to exclude inadvertent PCR mutations. The HisF-CC and HisF-C\*C proteins, which carry a His-6 tag at their C termini, were produced in *Escherichia coli* BL21(DE3). They were found in the insoluble fraction of the cell extract and solubilized, stepwise refolded, and purified as described for HisF-C (15).

**Cloning and Heterologous Expression of hisAF and hisFA and Purification of the Protein Products.** The chimeric genes *hisAF* and *hisFA* were cloned with and without a linker that encodes a Gly-Ser-Gly stretch. For construction of *hisAF* with linker, the N-terminal half of *hisA* (*hisA*-N) was amplified by PCR using the plasmid SK+/III P-P (14), and the oligonucleotide [1] 5'-AGC CAT ATG CTC GTT GTC CCG GCG ATA GAT-3' with a *Nde*I site (in bold) as the 5' primer and the oligonucleotide 5'-GGC GGA TCC ATC GAT TTC TCT CAG GGA TTT-3' with a *Bam*HI site (in bold) as the 3' primer. The amplified *hisA*-N was cloned into pET24a(+)-*hisF*-C<sub>2</sub> (see cloning of *hisF*-CC) by using *Nde*I and *Bam*HI, yielding pET24a(+)-*hisAlinkF*. The linker was removed by PCR with pET24a(+)-*hisAlinkF* as the template as follows. First, *hisA*-N was amplified by using the oligonucleotides [1] as the 5' primer and 5'-CAC GAC AAC GGC CTG ATC GAT TTC TCT CAG-3' as the 3' primer. The amplification product was used in a second PCR as the 5' primer together with the oligonucleotide 5'-GTG GGA TCC TTA CAA CCC CTC CAG TCT CAC GTT-3' with a *Bam*HI site (in bold) as the 3' primer. The resulting fragment was cloned into pET24a(+) by using *Nde*I and *Bam*HI, yielding pET24a(+)-*hisAF*. For construction of *hisFA* with linker, the C-terminal half of *hisA* (*hisA*-C) was amplified by PCR using the plasmid SK+/III P-P as the template, the oligonucleotide 5'-ATA GGA TCC GGT GTG GAG CCC GTG-3' with a *Bam*HI site (in bold) as the 5' primer, and the oligonucleotide [2] 5'-ATA GCG GCC GCG CGA GCA TAT CTC TTC ATC AC-3' with a *Not*I site (in bold) as the 3' primer. The amplified *hisA*-C was cloned into pET24a(+) by using *Bam*HI and *Not*I, yielding the plasmid pET24a(+)-*hisA*-C. Then, *hisF*-N was amplified by PCR using the plasmid SK+/III P-P as the template, the oligonucleotide [3] 5'-AGC CAT ATG CTC GCT AAA AGA ATA ATC GCG-3' with a *Nde*I site (in bold) as the 5' primer, and the oligonucleotide 5'-GCC GGA TCC ACT CCC AAA AGT TTG-3' with a *Bam*HI site (in bold) as the 3' primer. The amplified *hisF*-N was cloned into pET24a(+)-*hisA*-C by using *Nde*I and *Bam*HI, yielding the plasmid pET24a(+)-*hisFlinkA*. The linker was removed by PCR with pET24a(+)-*hisFlinkA* as the template as follows. The N-terminal half of *hisF* was amplified by using the oligonucleotide [3] as the 5' primer and 5'-GAA CAC GGG CTC CAC ACT CCC AAA AGT TTG-3' as the 3' primer. The amplification product was used in a second PCR as a 5' primer together with the oligonucleotide [2] as the 3' primer. The resulting fragment was cloned into pET24a(+) by using *Nde*I and *Not*I, yielding pET24a(+)-*hisFA*. All inserts were entirely sequenced to exclude inadvertent PCR mutations. Upon expression from these plasmids, no His-6 tag was attached to HisAF, whereas a His-6 tag was attached to the C terminus of HisFA.

Heterologous expression of *hisAF* and *hisFA* was conducted in *E. coli* BL21(DE3)c+ cells. HisAF was found in the soluble (≈10%) as well as in the insoluble (≈90%) fraction of the cell extract, but only protein from the insoluble fraction was purified. HisFA was found solely in the insoluble fraction. For further



**Fig. 2.** Design and spectroscopic characterization of HisF-C, HisF-CC, and HisF-C\*C. (A) Optimizing the interface between the half-barrels of HisF-CC yields HisF-C\*C. Superposition of the  $\beta$ -strands and the side chains of layer 4 from HisF (green) on the model of layer 4 of HisF-CC (yellow). (The figure was prepared with SWISS PDB VIEWER.) Within layer 4 of HisF-CC, alanine 124 was replaced by arginine (A124R), and alanine 220 was replaced by lysine (A220K) in the first HisF-C unit, yielding HisF-C\*C (see Fig. 1C). The side chain of the introduced K220 superimposes on K99 of HisF. In contrast, the side chain of the introduced R124 (which is located at the flexible N terminus) is pointing to the exterior of the barrel but might mimic the native R5 by moving inwards to form a salt-bridge cluster similar to that present in HisF. (B–D) HisF-C, HisF-CC, and HisF-C\*C have well defined secondary and tertiary structures. (B) Far-UV CD spectra. Protein concentrations are between 0.14 and 0.20 mg/ml,  $d = 0.1$  cm. The shown spectra are the mean of 10 individual spectra. (C) Fluorescence emission spectra (excitation at 280 nm). Protein concentrations are between 0.10 and 0.16 mg/ml. The emission maximum of HisF is at 323 nm, and the maxima of the other variants are at 333 nm. All spectra were recorded in 50 mM potassium phosphate (pH 7.5) at 23°C. (D) Near-UV CD spectra. Protein concentrations are as follows: HisF and HisF-C, 0.17 mg/ml,  $d = 5$  cm; HisF-C\*C, 3.0 mg/ml,  $d = 0.5$  cm; HisF-CC, 0.28 mg/ml,  $d = 1.0$  cm. The shown spectra are the mean of 10 individual spectra.

purification, both proteins were solubilized and stepwise refolded by dialysis against 50 mM potassium phosphate (pH 7.5) (15). The HisFA protein was pure to at least 95%, which was sufficient for its characterization. HisAF, which was pure to only  $\approx 90$ –95%, was loaded onto a HiLoad 26/60 Superdex 75 (320 ml, Amersham Pharmacia) that was equilibrated with 50 mM potassium phosphate and 300 mM KCl (pH 7.5). Elution was performed in the same buffer at a flow rate of 0.5 ml/min. Fractions with the highest protein content were pooled and dialyzed extensively against 50 mM potassium phosphate (pH 7.5). The characterization of HisAF and HisFA showed that the presence of the Gly–Ser–Gly linker did not make any measurable difference. For this reason, only data for HisAF and HisFA without linker are presented.

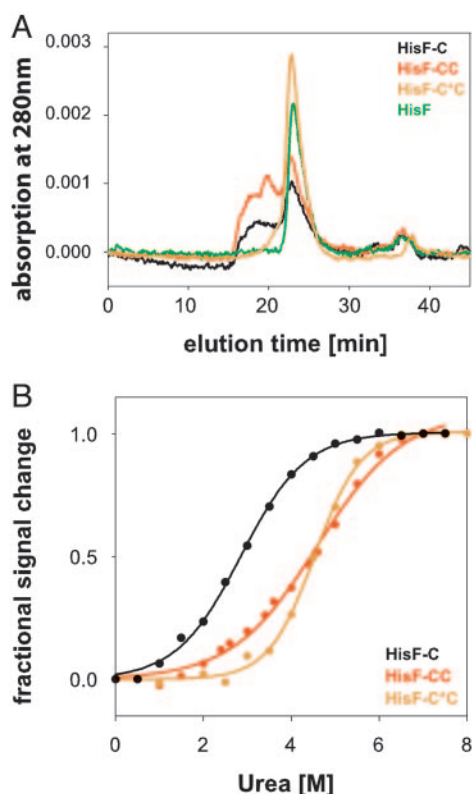
**Analytical Methods.** SDS/PAGE, protein concentration measurements, and CD spectroscopy were carried out as described in ref. 15. Analytical gel filtration was performed by using a calibrated Superdex 75 column (Amersham Pharmacia). The proteins (0.03 mg for HisF-C, HisF-CC, and HisF-C\*C; 0.24 mg for HisAF and HisFA) were eluted at a flow rate of 0.5 ml/min in 50 mM potassium phosphate and 300 mM KCl (pH 7.5) at 23°C. Fluorescence spectroscopy was performed with a Cary Eclipse spectrophotometer (Varian). Protein unfolding induced by urea or guanidinium chloride was followed by the decrease of the CD signal at

222 nm, or of the fluorescence emission at 320 nm (HisF, HisF-C, HisF-CC, HisF-C\*C, and HisAF), 322 nm (HisFA), or 340 nm (HisA) after excitation at 280 nm. The proteins were incubated with different concentrations of the chaotropic agents, and the signals were detected after different time intervals until no further change was observed. After complete unfolding in 8 M urea or 6 M guanidinium chloride, all proteins could be refolded by removing the chaotropic agent by means of dilution or dialysis, by using 50 mM potassium phosphate buffer (pH 7.5). Limited proteolysis was performed in 50 mM Tris (pH 8.0) at 25°C, containing 0.2  $\mu$ M trypsin and 10  $\mu$ M HisAF. The reaction was stopped after different time intervals by adding SDS/PAGE sample buffer and heating for 5 min at 95°C. The time course of proteolysis was followed on Tris-*N*-tris(hydroxymethyl)methylglycine gels containing 20% acrylamide (21).

## Results and Discussion

**Production of a Stable ( $\beta\alpha$ )<sub>8</sub>-Barrel from Two HisF-C Half-Barrels.** Two copies of the *hisF*-C gene from *T. maritima* (15), each encoding the sequence from  $\beta$ -strand 5 through  $\alpha$ -helix 8 (residues 123–253), were cloned in tandem and connected by a Gly–Ser–Gly linker, yielding HisF-CC (Fig. 1C). The amino acid side chains that point to the interior of the wild-type HisF barrel form four optimally packed layers that lie on top of each other and are numbered from the C-terminal to the N-terminal face (22). Each

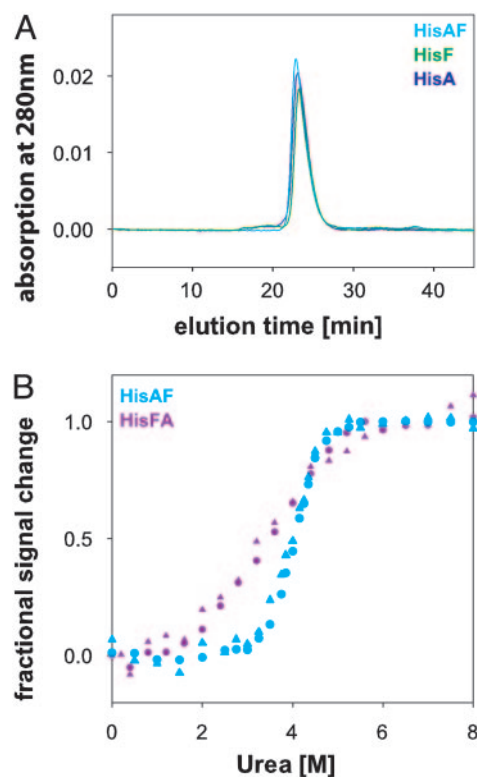




**Fig. 3.** Association states and stabilities of HisF-C, HisF-CC, and HisF-C\*C. (A) Analytical gel filtration reveals different association states. The main peaks correspond to a molecular mass of 29.7 kDa, which is equivalent to a monomer for HisF, HisF-C\*C, and HisF-CC, and to a homodimer for HisF-C. The faster eluting peaks correspond to higher association states of HisF-C and HisF-CC, which are not well defined. (B) Urea-induced denaturation shows that HisF-C\*C and HisF-CC are more stable than HisF-C and that HisF-C\*C unfolds with a higher cooperativity than HisF-CC. Proteins at a concentration between 0.11 and 0.13 mg/ml were incubated with the given concentrations of urea in 50 mM potassium phosphate (pH 7.5) at 23°C. Unfolding was followed by recording the decrease of the fluorescence emission at 320 nm after excitation at 280 nm. The transition midpoint of HisF-C is at 2.8 M urea, and the transition midpoints of HisF-CC and HisF-C\*C are at 4.4 M urea. The lines that connect the individual points were drawn as a visual aid.

layer is either formed by the side chains from residues of the four odd- or even-numbered  $\beta$ -strands, leading to a fourfold symmetric layer arrangement. The fourth layer differs from the other three layers in that it is formed by the four charged residues arginine 5 ( $\beta$ -strand 1), glutamate 46 ( $\beta$ -strand 2), lysine 99 ( $\beta$ -strand 4), and glutamate 167 ( $\beta$ -strand 6), whereas the other layers are formed by hydrophobic and polar side chains. Moreover, arginine 5 does not fit into the fourfold symmetry pattern but locates the guanidinium group of its long side chain in front of the small side chain of alanine 220 of  $\beta$ -strand 8, which is the regular member of layer 4. Modeling of the putative HisF-CC barrel by superposition of a second HisF-C half onto the N-terminal half of HisF suggested that the residues in layers 1–3 are well packed. Within layer 4, however, HisF-CC cannot form the conserved and putatively stabilizing salt-bridge cluster that is found in HisF, because of the presence of alanine 124 ( $\beta$ -strand 5) and alanine 220 ( $\beta$ -strand 8) equivalent to arginine 5 ( $\beta$ -strand 1) and lysine 99 ( $\beta$ -strand 4). To reconstitute the salt-bridge cluster, alanine residues 124 and 220 in the N-terminal half of HisF-CC were replaced by arginine and lysine, respectively, yielding HisF-C\**C* (Figs. 1C and 2A).

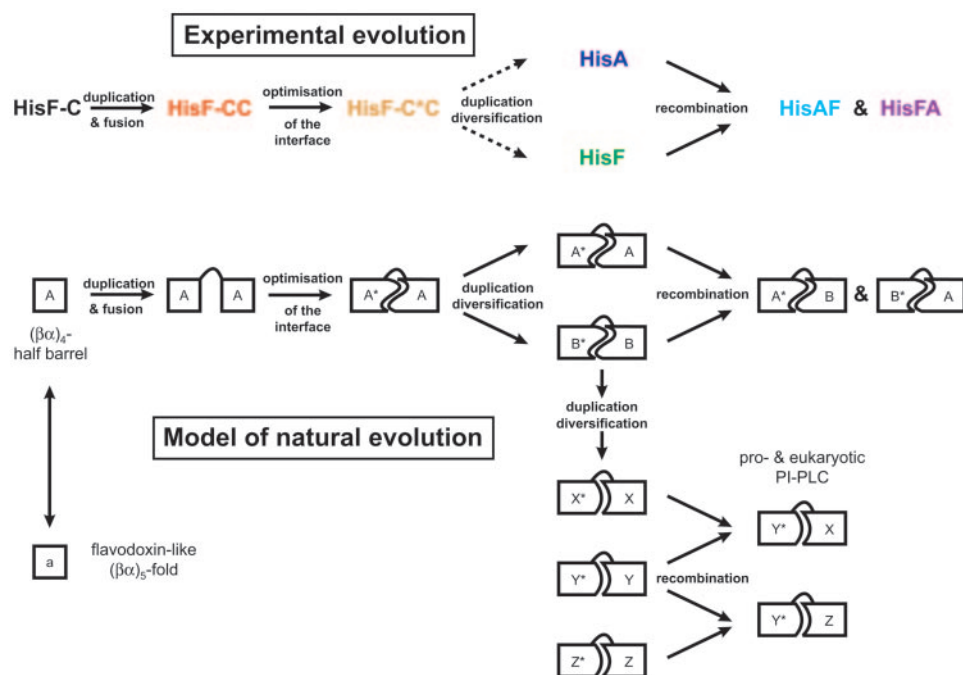
HisF-C, HisF-CC, and HisF-C\**C* were produced in *E. coli*, purified, and characterized in comparison with wild-type HisF.



**Fig. 4.** Association states (*A*) and stabilities (*B*) of HisAF and HisFA. (*A*) Analytical gel filtration. The main peaks of HisAF, HisF, and HisA correspond to molecular masses of 29.5, 28.1, and 26.7 kDa, respectively, which are equivalent to monomers. (*B*) Urea-induced denaturation shows that HisAF is more stable than HisFA. Proteins at concentrations of 0.1 mg/ml were incubated with the given concentrations of urea in 50 mM potassium phosphate (pH 7.5) at 23°C. The loss of the tertiary structure was followed by recording the decrease of the fluorescence emission (circles) at 320 nm (HisAF) or 322 nm (HisFA) after excitation at 280 nm. The loss of the secondary structure was recorded by the decrease of the CD signal at 222 nm (triangles).

The far-UV CD spectra of the three variants were virtually identical and similar to that of HisF (Fig. 2*B*), suggesting that all proteins have well defined secondary structures (23). HisF-C, HisF-CC, HisF-C\*C, and HisF contain the same single tryptophan residue 156 in  $\alpha$ -helix 5 (10) (Fig. 1*A* and *B*), which allows one to compare their tertiary structures on the basis of fluorescence and near-UV CD spectroscopy. The fluorescence emission spectra of HisF-C, HisF-CC, and HisF-C\*C lie at 333 nm, in comparison with 323 nm for HisF (Fig. 2*C*). Moreover, their distinct near-UV CD spectra are practically identical and similar to that of HisF (Fig. 2*D*). These results suggest that the indole chromophore of tryptophan 156 is in a comparably asymmetric environment in the three variants and that it is almost as well shielded from solvent as in HisF (23).

In analytical gel filtration experiments, a significant fraction of purified HisF-C and HisF-CC correspond to ill-defined oligomers (Fig. 3A). The main peaks of HisF-C and HisF-CC however, elute at a molecular mass of 29.7 kDa, corresponding to dimeric HisF-C and monomeric HisF-CC. In contrast to HisF-CC, HisF-C\*C is solely monomeric, suggesting that the substitutions in layer 4 (Fig. 2A) stabilize a more compact structure. The reversible unfolding by urea revealed that, at identical protein concentrations, HisF-CC is considerably more stable than HisF-C (Fig. 3B). HisF-C\*C is as stable as HisF-CC but unfolds with a higher cooperativity, testifying to a more compact structure, in accordance with the sharp elution peak observed in the analytical gel filtration (Fig. 3A). Finally,



**Fig. 5.** Evolving  $(\beta\alpha)_8$ -barrels by duplicating and fusing, and mixing and matching  $(\beta\alpha)_4$ -half-barrels. Model of the natural evolution of  $(\beta\alpha)_8$ -barrel enzymes from half-barrels (Lower) and its experimental reconstruction in this study (Upper). The variants generated and characterized are color-coded as in Figs. 2–4. The primordial  $(\beta\alpha)_4$ -half-barrel was mimicked by HisF-C. The duplication of its gene and fusion yielded the gene for HisF-CC, and the subsequent optimization of the interface between the two identical halves resulted in HisF-C\*C. HisF-C\*C mimics an ancestral  $(\beta\alpha)_8$ -barrel, from which HisA and HisF might have evolved by means of further gene duplication and diversification events. Recombining  $(\beta\alpha)_4$ -half-barrels, mimicked by HisA-N, HisF-N, HisA-C, and HisF-C, leads to new  $(\beta\alpha)_8$ -barrels, mimicked by HisAF and HisFA. Through further steps of duplication and diversification, a repertoire of less symmetrical  $(\beta\alpha)_8$ -barrel enzymes evolved, which was extended by recombination. An example is provided by the prokaryotic and eukaryotic phospholipases C (30) (see text for details). A comprehensive database search suggests an evolutionary linkage between  $(\beta\alpha)_4$ -half-barrels and the  $(\beta\alpha)_5$ -flavodoxin-like fold family (19).

HisF-C, HisF-CC, and HisF-C\*C were tested for catalysis of the HisF reaction under steady-state conditions. None of them showed measurable activity, even at protein concentrations of 20  $\mu$ M.

**Production of Chimeric  $(\beta\alpha)_8$ -Barrels by the Crosswise Fusion of  $(\beta\alpha)_4$  Half-Barrels from HisA and HisF.** The two chimeric proteins HisAF and HisFA were produced in *E. coli*, purified, and characterized in comparison with their parent proteins, HisA and HisF. In HisAF,  $(\beta\alpha)_{1-4}$  is derived from HisA, and  $(\beta\alpha)_{5-8}$  is derived from HisF. In the mirror chimera HisFA,  $(\beta\alpha)_{1-4}$  is derived from HisF, and  $(\beta\alpha)_{5-8}$  is derived from HisA (Fig. 1C).

The far-UV CD spectrum of HisAF (see Fig. 6A, which is published as supporting information on the PNAS web site) testifies to a well defined secondary structure. HisAF and HisF contain the same single tryptophan residue 156 in helix  $\alpha 5$  (Fig. 1). The near-UV CD and the fluorescence emission spectra of HisAF and HisF are almost identical (Fig. 6B and C), suggesting that the indole chromophore is comparably well shielded in the asymmetric interior of the two proteins. Analytical gel filtration showed that HisAF elutes at the same time and as an equally sharp peak as HisA and HisF, proving that it is a homogenous monomer (Fig. 4A). To assess the conformational stability of HisAF, its reversible unfolding was induced by urea and followed by far-UV CD and fluorescence spectroscopy (Fig. 4B). The two equilibrium-unfolding traces superimpose well, which indicates that the secondary and tertiary structures are lost simultaneously. Moreover, unfolding is highly cooperative with a transition midpoint at  $\approx 4$  M urea. These data show that HisAF has native-like properties and is comparably stable as an average natural protein (24). HisA and HisF cannot be unfolded completely by urea (data not shown). To compare their stabilities

with that of HisAF, all three proteins were denatured by guanidinium chloride, and unfolding was followed by fluorescence spectroscopy. The transition midpoints of HisA, HisF, and HisAF occurred at 3.5, 3.1, and 1.8 M guanidinium chloride, showing that the chimera is less stable against denaturant than its parent proteins (see Fig. 7, which is published as supporting information on the PNAS web site). Limited proteolysis was applied to further compare the stabilities of HisA and HisAF. Both proteins were cleaved by trypsin at a similar rate (data not shown) after the same arginine residue, which is located in the flexible loop that connects  $\beta$ -strand 1 with  $\alpha$ -helix 1 (Fig. 1B). HisAF was tested for catalysis of the HisA and the HisF reaction under steady-state conditions. It displayed neither activity, even at a concentration of 20  $\mu$ M. Accordingly, the plasmid-encoded *hisAF* gene was unable to complement on selective medium auxotrophic *E. coli* strains that lack a functional *hisA* or *hisF* gene on their chromosome (data not shown). We conclude that HisAF is catalytically inactive, both *in vitro* and *in vivo*.

According to far-UV CD spectroscopy, HisFA adopts a well defined secondary structure (Fig. 6A). HisA and HisFA contain the same single tryptophan residue 138 in the loop between  $\beta$ -strand 5 and  $\alpha$ -helix 5 (Fig. 1). However, the near-UV CD spectrum of HisFA is less pronounced than that of HisA, suggesting that the environment of the indole chromophore is less asymmetric and that its tertiary structure is less well ordered (Fig. 6B). In contrast, the fluorescence emission maximum of HisFA is found at a lower wavelength (341 nm) than that of HisA (347 nm) (Fig. 6C), suggesting that tryptophan 138 is less solvent-exposed in HisFA than in HisA. Analytical gel filtration shows that HisFA forms a mixture of various association states, which are not well defined (data not shown). The unfolding of HisFA occurs at lower concentrations of urea and is less

cooperative compared with HisAF (Fig. 4B). These results demonstrate that HisFA is less stable and has a less-defined and less compact structure than HisAF, in accordance with its weakly pronounced near-UV CD spectrum. Along the same lines, incubation with trypsin results in the complete degradation of HisFA within 5 min (data not shown).

**Implications for the Evolution of  $(\beta\alpha)_8$ -Barrels.** It has been postulated that HisA and HisF evolved from a common  $(\beta\alpha)_4$ -half-barrel by a series of gene duplication and diversification events (4, 11). In the present work, the first steps of this hypothetical evolutionary pathway were reconstructed experimentally, by using HisF-C as a model for the ancestral half-barrel (Fig. 5). The tandem fusion of two HisF-C units led to HisF-CC, which is more stable than HisF-C (Fig. 3B). In accordance with this finding, it has been postulated that protein fusion in general increases the conformational stability of the linked proteins, because it decreases the translational and rotational entropy of the unfolded state (25–27). Along these lines, fusion of the subunits of the homodimeric gene V protein of bacteriophage f1 resulted in an increased stability and folding rate (28). The HisF-CC protein did not form a homogeneous monomer but still had a tendency to aggregate (Fig. 3A), probably because steric hindrance and/or electrostatic repulsion prevented the proper association of the two identical half-barrels. To optimize this association, the native salt-bridge cluster of layer 4 was reconstituted within HisF-CC, which yielded the homogeneous and compact protein HisF-C\**C*. The improved stability and solubility of HisF-C\**C* compared with HisF-CC has been anticipated by the “Rosetta stone model” for the evolution of protein–protein interactions (29). This model postulates that protein fusion leads to unspecific interactions between the linked domains, which then are optimized by interface mutations. The model outlined in Fig. 5 also suggests that  $(\beta\alpha)_4$ -half-barrels, although an

integral part of  $(\beta\alpha)_8$ -barrels, are independently evolving domains that were mixed and matched to other  $(\beta\alpha)_4$ -half-barrels or quite different domains in the course of evolution (18). In support of this idea, the chimeric proteins HisAF and HisFA could be isolated in pure form and characterized. Whereas HisFA is relatively insoluble, labile, and forms ill-defined oligomers, HisAF is highly soluble, stable, and monomeric. These findings indicate that the  $(\beta\alpha)_8$ -barrels HisA and HisF are indeed composed of interchangeable  $(\beta\alpha)_4$  units. In other  $(\beta\alpha)_8$ -barrels, such  $(\beta\alpha)_4$ -half-barrels might have distinct functional properties, the combination of which with different protein domains would result in a remarkable increase of catalytic versatility. An example is provided by the phosphoinositide-specific phospholipases C. Whereas the structures of the C-terminal halves of eukaryotic and prokaryotic phosphoinositide-specific phospholipases C are unrelated, their N-terminal  $(\beta\alpha)_4$ -half-barrels, which contain all catalytically essential residues, superimpose with an rms deviation of only 1.85 Å for 104 equivalent C $\alpha$  atoms (30). Obviously, the catalytic core domain of phosphoinositide-specific phospholipases C is a  $(\beta\alpha)_4$ -half-barrel, the structural and functional properties of which are modified by the different domains that are fused to its C terminus.

In summary, we were able to reconstruct experimentally putative events in the course of  $(\beta\alpha)_8$ -barrel evolution, generating the stable and monomeric HisF-C\**C* and HisAF proteins. These proteins have well defined tertiary structures and, therefore, provide an appropriate scaffold for the establishment of catalytic activity, which is the prerequisite for an evolutionary advantageous function.

We thank Drs. Kasper Kirschner and Franz X. Schmid for valuable comments on the manuscript. This work was supported by Deutsche Forschungsgemeinschaft Grant STE 891/4-1, 4-2.

1. Fani, R., Mori, E., Tamburini, E. & Lazcano, A. (1998) *Origins Life Evol. Biosphere* **28**, 555–570.
2. Lynch, M. & Conery, J. S. (2000) *Science* **290**, 1151–1155.
3. Wierenga, R. K. (2001) *FEBS Lett.* **492**, 193–198.
4. Höcker, B., Jürgens, C., Wilmanns, M. & Sterner, R. (2001) *Curr. Opin. Biotechnol.* **12**, 376–381.
5. Vega, M. C., Lorentzen, E., Linden, A. & Wilmanns, M. (2003) *Curr. Opin. Chem. Biol.* **7**, 694–701.
6. Copley, R. R. & Bork, P. (2000) *J. Mol. Biol.* **303**, 627–641.
7. Nagano, N., Orenge, C. A. & Thornton, J. M. (2002) *J. Mol. Biol.* **321**, 741–765.
8. Gerlt, J. A. & Raushel, F. M. (2003) *Curr. Opin. Chem. Biol.* **7**, 252–264.
9. Wise, E. L. & Rayment, I. (2004) *Acc. Chem. Res.* **37**, 149–158.
10. Lang, D., Thoma, R., Henn-Sax, M., Sterner, R. & Wilmanns, M. (2000) *Science* **289**, 1546–1550.
11. Henn-Sax, M., Höcker, B., Wilmanns, M. & Sterner, R. (2001) *Biol. Chem.* **382**, 1315–1320.
12. Jürgens, C., Strom, A., Wegener, D., Hettwer, S., Wilmanns, M. & Sterner, R. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 9925–9930.
13. Leopoldseder, S., Claren, J., Jürgens, C. & Sterner, R. (2004) *J. Mol. Biol.* **337**, 871–879.
14. Thoma, R., Schwander, M., Liebl, W., Kirschner, K. & Sterner, R. (1998) *Extremophiles* **2**, 379–389.
15. Höcker, B., Beismann-Driemeyer, S., Hettwer, S., Lustig, A. & Sterner, R. (2001) *Nat. Struct. Biol.* **8**, 32–36.
16. Beismann-Driemeyer, S. & Sterner, R. (2001) *J. Biol. Chem.* **276**, 20387–20396.
17. Henn-Sax, M., Thoma, R., Schmidt, S., Hennig, M., Kirschner, K. & Sterner, R. (2002) *Biochemistry* **41**, 12032–12042.
18. Gerlt, J. A. & Babbitt, P. C. (2001) *Nat. Struct. Biol.* **8**, 5–7.
19. Höcker, B., Schmidt, S. & Sterner, R. (2002) *FEBS Lett.* **510**, 133–135.
20. Sarkar, G. & Sommer, S. S. (1990) *BioTechniques* **8**, 404–407.
21. Schagger, H. & von Jagow, G. (1987) *Anal. Biochem.* **166**, 368–379.
22. Douangamath, A., Walker, M., Beismann-Driemeyer, S., Vega-Fernandez, M. C., Sterner, R. & Wilmanns, M. (2002) *Structure (London)* **10**, 185–193.
23. Schmid, F. X. (1997) in *Protein Structure: A Practical Approach*, ed. Creighton, T. E. (Oxford Univ. Press, New York), pp. 259–295.
24. Pfeil, W. (2001) *Protein Stability and Folding: A Collection of Thermodynamic Data* (Springer, Berlin).
25. Erickson, H. P. (1989) *J. Mol. Biol.* **206**, 465–474.
26. Terwilliger, T. C. (1995) *Adv. Protein Chem.* **46**, 177–215.
27. Nagi, A. D. & Regan, L. (1997) *Folding Des.* **2**, 67–75.
28. Liang, H., Sandberg, W. S. & Terwilliger, T. C. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 7010–7014.
29. Marcotte, E. M., Pellegrini, M., Ng, H. L., Rice, D. W., Yeates, T. O. & Eisenberg, D. (1999) *Science* **285**, 751–753.
30. Heinz, D. W., Essen, L. O. & Williams, R. L. (1998) *J. Mol. Biol.* **275**, 635–650.